Human-centered Explainable AI: Towards a Reflective Sociotechnical Approach

Year: 2020 | Citations: 253 | Authors: Upol Ehsan, Mark O. Riedl

---

Abstract

Explanations--a form of post-hoc interpretability--play an instrumental role in making systems accessible as AI continues to proliferate complex and sensitive sociotechnical systems. In this paper, we introduce Human-centered Explainable AI (HCXAI) as an approach that puts the human at the center of technology design. It develops a holistic understanding of "who" the human is by considering the interplay of values, interpersonal dynamics, and the socially situated nature of AI systems. In particular, we advocate for a reflective sociotechnical approach. We illustrate HCXAI through a case study of an explanation system for non-technical end-users that shows how technical advancements and the understanding of human factors co-evolve. Building on the case study, we lay out open research questions pertaining to further refining our understanding of "who" the human is and extending beyond 1-to-1 human-computer interactions. Finally, we propose that a reflective HCXAI paradigm-mediated through the perspective of Critical Technical Practice and supplemented with strategies from HCI, such as value-sensitive design and participatory design--not only helps us understand our intellectual blind spots, but it can also open up new design and research spaces.